Leading Edge
Review
Cell

Alternative Splicing: New Insights from Global Analyses

Benjamin J. Blencowe[1,*]
[1]Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, Centre for Cellular and Biomolecular Research, Donnelly CCBR Building, University of Toronto, 160 College Street, Room 1016, Toronto, ON M5S 3E1, Canada
*Contact: b.blencowe@utoronto.ca
DOI 10.1016/j.cell.2006.06.023

Recent analyses of sequence and microarray data have suggested that alternative splicing plays a major role in the generation of proteomic and functional diversity in metazoan organisms. Efforts are now being directed at establishing the full repertoire of functionally relevant transcript variants generated by alternative splicing, the specific roles of such variants in normal and disease physiology, and how alternative splicing is coordinated on a global level to achieve cell- and tissue-specific functions. Recent progress in these areas is summarized in this review.

One of the most remarkable observations stemming from the sequencing of genomes of diverse species is that the number of protein-coding genes in an organism does not correlate with its overall cellular complexity. For example, *Drosophila melanogaster* has fewer protein-coding genes than the nematode *Caenorhabditis elegans* (~14,000 versus ~19,000). Meanwhile, mammalian species have similar numbers of protein-coding genes as *Arabidopsis thaliana* (~20,000–25,000) and only four times the number found in the budding yeast *Saccharomyces cerevisiae* (~6,000). These observations indicate that mechanisms acting to regulate and diversify gene functions must have played a major role in the evolution of specialized cell types and activities that are typically associated with complex metazoans. Alternative splicing (AS), the process by which the exons of primary transcripts (pre-mRNAs) from genes can be spliced in different arrangements to produce structurally and functionally distinct mRNA and protein variants, may be one of the most extensively used mechanisms that accounts for the greater macromolecular and cellular complexity of higher eukaryotic organisms.

AS has numerous critical roles in metazoan organisms (Black, 2003; Matlin et al., 2005). Despite many focused studies on the functions and mechanisms of AS that are associated with specific transcripts, high-throughput experimental approaches for systematically elucidating the roles of AS events are only now beginning to be used. Considerable effort has been directed at the genome-wide identification of AS events in different cell and tissue types and under different conditions in order to establish the extent of functionally relevant AS events. Initial analyses of the resulting data sets are revealing important global regulatory features of AS.

Technologies for the Global Analysis of AS

The availability of sequenced genomes and large databases of sequenced transcripts, primarily comprising expressed sequence tags (ESTs) and smaller numbers of cDNA sequences, has provided a rich source of information for the identification and analysis of AS events. EST and cDNA sequences can be aligned to genomic sequences using programs that search for conserved splice-site consensus sequences adjacent to the gaps created by intron sequences between the aligned exons. Contigs of genomic exons extracted in this manner are realigned to the corresponding ESTs and cDNAs such that clusters of aligned transcripts with or without middle exon alignments (indicative of an AS event) can be systematically identified. Large databases of AS events mined in this manner have been established for several species, including human, mouse, and rat (Modrek and Lee, 2002; Lee et al., 2003; Thanaraj et al., 2004; Zheng et al., 2005). However, a major limitation of AS analyses employing transcript sequence data is that EST coverage is typically biased toward the 3′ and 5′ ends of transcripts, and in general there are insufficient numbers of sequenced transcripts to infer the frequency with which specific alternative exons are included or skipped in a given cell or tissue source or under particular experimental conditions (Johnson et al., 2003; Pan et al., 2004).

Some of the limitations inherent in the analysis of EST/cDNA have been overcome by the development of custom microarrays and computational tools, as well as differential hybridization techniques, that permit the large-scale profiling of AS (Yeakley et al., 2002; Johnson et al., 2003; Wang et al., 2003; Le et al., 2004; Pan et al., 2004; Stolc et al., 2004; Watahiki et al., 2004; Blanchette et al., 2005; Ule et al., 2005; Shai et al., 2006; Sugnet et al., 2006). Several of the AS microarrays that have been described contain thousands

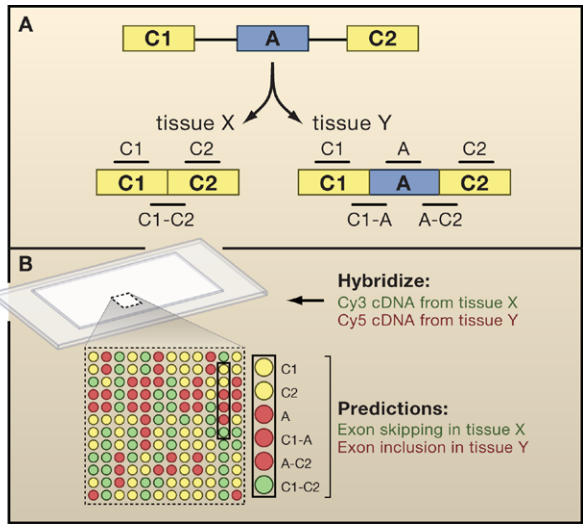

Figure 1. Microarray Profiling of AS

(A) Cassette-type AS events, which consist of an alternative exon "A" (blue box) flanked by intron sequences and constitutively spliced exons "C1" and "C2" (yellow boxes), represent the most common type of regulated AS event and have been the most extensively analyzed by sequence- and microarray-based methods (see also Figure 2). In the example shown, a set of six probes, three targeted to exons (C1, A, and C2 probes) and three to splice-junction sequences (C1-A, A-C2, and C1-C2), permits the quantitative profiling of a cassette alternative exon between two different tissue types (tissues X and Y). In the example shown, the cassette alternative exon is skipped in tissue X and included in tissue Y. Adapted from Pan et al. (2004).

(B) Hypothetical hybridization pattern obtained from microarray profiling of AS using Cy3-labeled tissue X cDNA (green) and Cy5-labeled tissue Y cDNA (red). Green and red spots indicate detection of signals from probes hybridized to the labeled cDNA from tissues X and Y, respectively. Yellow spots indicate detection of signal from exons or splice junctions that are expressed in both tissues X and Y. Data processed using a suitable algorithm from this microarray format can permit the accurate prediction of AS levels for thousands of cassette-type alternative exons (Shai et al., 2006).

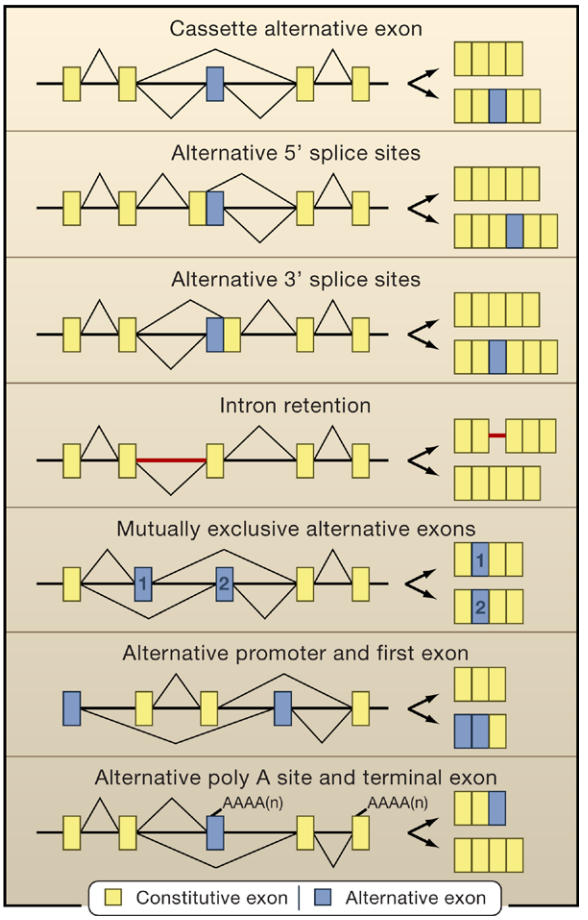

Figure 2. AS Events in Metazoan Transcripts

Types of AS that are responsible for the generation of functionally distinct transcripts are depicted. Blue boxes indicate alternative exons.

of sets of anchored oligonucleotide probes on glass slides, with each set typically combining probes that are specific to individual exons and/or splice-junction sequences formed by inclusion or skipping of exons (Figure 1). This type of format has permitted the discovery of new AS events not previously detected in cDNA or EST sequences (Johnson et al., 2003) and the large-scale detection of cell- and tissue-specific AS events involving exons that were initially identified using EST/cDNA sequence data (Pan et al., 2004). More recently, this microarray format has facilitated the global analysis of alternative exons regulated by specific splicing factors (Blanchette et al., 2005; Ule et al., 2005) and has led to the discovery of sequence motifs that correlate with tissue-specific AS (Sugnet et al., 2006). Another microarray format employing a fiber-optic-based system for the detection of specific splice variants has been described, and this approach has been used to monitor splice variants in different transformed cell lines and tumors (Yeakley et al., 2002; Li et al., 2006; Zhang et al., 2006).

Insights into the Global Complexity of AS

Based on systematic analyses of ESTs and cDNAs, especially from mammalian species, it is apparent that the most common type of AS, accounting for at least one-third of known AS events, involves cassette-type alternative exons. These exons, which are either skipped or included in the final message, are flanked by intron sequences (Figure 1 and Figure 2). Alternative selection of 5′ or 3′ splice sites within exon sequences are also frequent, together accounting for at least one-quarter of the known AS events (Figure 2). This type of AS is capable of introducing subtle changes into coding sequences, differing by as little as a single codon. For example, approximately 30% of human genes contain NAGNAG sequences at the 3′ ends of introns, which have the potential to act as tandem splice-site acceptors (Hiller et al., 2004). However, it is not clear to what extent such subtle variants are functionally significant, and a recent analysis suggests that a large fraction may arise as a consequence of stochastic binding of the spliceosome at neighboring splice sites (Chern et al., 2006)

(see below). Other types of AS events include retained introns (Ohler et al., 2005) and exons that are spliced in a mutually exclusive fashion (Figure 2).

In addition to the AS mechanisms mentioned above, the exon composition of transcripts is often altered by differential selection of transcription initiation and 3′ end processing/termination sites (Figure 2), and these events can impact on adjacent or distal AS events in the same transcript (Zavolan et al., 2003; Kornblihtt, 2005). In several studies, it has been shown that transcription factors acting at the level of initiation and elongation can impact splice-site selection (Kornblihtt, 2006). In particular, factors resulting in reduced rates of RNA polymerase II (Pol II) elongation can increase the inclusion of alternative exons. One model that has been proposed to explain this effect is that reducing the rate of Pol II elongation kinetically favors the recognition of relatively weak splicing signals surrounding an alternative exon over the inherently stronger splicing signals that otherwise favor splicing of the neighboring upstream and downstream constitutive exons, resulting in skipping of the alternative exon (Kornblihtt, 2006).

It is worth noting that for each type of AS event mentioned above, a distinction can be made between mechanisms controlling splice-site selection involving the binding of regulatory factors and mechanisms that may operate in a stochastic manner. For example, the selection of mutually exclusive exons 2 and 3 in the α-tropomyosin pre-mRNA involves the repression of exon 3 by *trans*-acting regulatory factors in smooth muscle cells (Gromak et al., 2003), whereas a stochastic process involving competing base-pairing interactions between intron sequences appears to play an important role in selection of mutually exclusive alternative exons in the *Dscam* gene of *Drosophila* (Graveley, 2005). Remarkably, this gene, which functions in both axon guidance and immune defense (Schmucker et al., 2000; Watson et al., 2005; Chen et al., 2006), has the potential to encode over 38,000 different splice variants, primarily via tandem arrays of mutually exclusive alternative exons, many of which overlap the coding regions for immunoglobulin-like domains. Finally, it should also be kept in mind that each of the types of AS summarized above and shown in Figure 2 can occur within both translated and untranslated regions (UTRs) of transcripts.

Consistent with the evidence that RNA Pol II can influence splice-site selection, experimental work indicates that most splicing events, as well as other steps in pre-mRNA processing, occur on pre-mRNA as it emerges from Pol II (Bentley, 2005). Recent evidence further indicates that the exons of nascent transcripts can be "tethered" to Pol II during pre-mRNA processing (Dye et al., 2006). In addition to potentially participating in AS, tethering of nascent exons to Pol II, together with the formation of networks of interactions across introns and exons (see below), may help to ensure order in the splicing process such that correct pairs of splice sites are united, even when separated by introns that are tens of thousands of bases in length (Ibrahim el et al., 2005; Dye et al., 2006). Downstream exons are rarely, if ever, inserted between upstream exons, and *trans*-splicing, although common in nematodes and trypanosomes, also rarely occurs in most metazoan species (Horiuchi and Aigaki, 2006). However, given the constraints imposed by the transcription and splicing machineries, AS can result in anywhere from a single skipping event per gene to up to tens of thousands of potential splice variants per gene.

It has been inferred from the analyses of EST and AS microarray data that over two-thirds of human genes and over 40% of *Drosophila* genes contain one or more alternative exons (Johnson et al., 2003; Stolc et al., 2004). In sharp contrast to these estimates, *S. cerevisiae* has only several known regulated splicing events and lacks orthologs for many factors associated with regulated splicing in metazoans, such as members of the serine/arginine-repeat (SR) and heterogeneous nuclear ribonucleoprotein (hnRNP) families of proteins (Davis et al., 2000; Boucher et al., 2001). The lack of sufficiently large data sets of AS microarray data and sequenced ESTs and cDNAs has prevented reliable estimates of the proportions of genes that undergo AS in other organisms (Brett et al., 2002; Kim et al., 2004). There are also insufficient data currently available to accurately assess the overall number of AS events in any one organism. In the case of human genes known to undergo AS, based on available transcript sequence and microarray profiling data, it is estimated that there are between one and two AS events per multi-intron gene (Lander et al., 2001; Johnson et al., 2003). However, this number may rise considerably as large-scale analyses of AS are performed in more specialized cell and tissue types. In addition, certain genes that undergo "extreme AS," such as the *Dscam* gene in *Drosophila*, are clearly outliers that confound the derivation of accurate estimates for the overall frequency and complexity of AS. Nevertheless, the results of recent bioinformatic and microarray-based analyses have provided several interesting insights into the nature of the evolutionary forces that can impact overall AS frequencies, as well as on additional global features of AS that will be discussed below.

Evolutionary Forces Acting on Global AS Frequencies

Several recent studies have provided evidence that the number of AS events per gene is inversely correlated with gene or paralog copy number, indicating that gene duplication, followed by divergence of paralog functions, may have reduced selection pressure to diversify gene functions by AS (Kopelman et al., 2005; Su et al., 2006). Moreover, sequence- and microarray-based analyses have indicated that AS events occur more often in transcripts from genes expressed in functionally complex tissues with diverse cell types, such as the brain and testis, or from genes expressed within individual cell types that have undergone selection to provide diverse functions, such as in the immune system (Modrek et al.,

2001; Johnson et al., 2003; Yeo et al., 2004; Watson et al., 2005). Indeed, the number of documented examples of regulated AS in the brain is vast, and many of these events are implicated in complex processes such as the control of synaptic plasticity associated with cognition and other neural processes (Lipscombe, 2005; Ule and Darnell, 2006). One example of particular interest involves AS of the apolipoprotein E receptor (*Apoer2*) gene, which is important for neuronal cell migration during brain development, as well as for long-term potentiation (LTP) in adult mice. Deletion of a single, activity-dependent alternative exon in the *Apoer2* gene prevents tyrosine phosphorylation of NMDA receptor subunits by Reelin, a ligand of Apoer2 (Beffert et al., 2005). Mice missing this *Apoer2* exon perform poorly in memory and learning tasks. Besides demonstrating a role for AS of *Apoer2* transcripts in LTP, this study also serves to illustrate the importance of targeting individual alternative exons when attempting to understand the regulation of complex biological processes.

Which AS Events Are Functionally Significant?

Independent sources of experimental data from AS microarray profiling, RT-PCR assays, and analyses of EST and cDNA sequences have provided consistent evidence that many alternatively spliced transcripts are low in abundance, especially those that are not conserved over the ~80 million year time interval separating human and mouse. In fact, comparisons of sequenced transcripts corresponding to large numbers of ortholog gene pairs in human and mouse indicate that only 10%–20% of cassette-type AS events are conserved between these species (Modrek and Lee, 2003; Sorek et al., 2004; Pan et al., 2005; Yeo et al., 2005). The remaining 80%–90% of human- and mouse-specific cassette-type AS events can be separated into two categories: (1) those involving relatively recently gained "genome-specific" exons, which tend to be included at low levels in spliced mRNA (also referred to as "minor-form" variants) (Lev-Maor et al., 2003; Modrek and Lee, 2003), and (2) those involving exons that are conserved but alternatively spliced in only one species, which are generally skipped at low levels in spliced mRNA (Pan et al., 2004, 2005).

The sequences of exons and flanking intron regions of AS events present in human and mouse are on average significantly more conserved than those of both species-specific AS and constitutive splicing events. Conserved AS events also more frequently preserve open reading frames (Philipps et al., 2004; Resch et al., 2004a; Sorek et al., 2004; Yeo et al., 2005) (see below). Remarkably, comparisons of nucleotide substitution rates in human and mouse indicate that, on average, as many as half of the synonymous positions in codons of conserved alternatively spliced exons are under selection pressure (Xing and Lee, 2005a). These and other interspecies comparisons support the conclusion that conserved alternative exons and their flanking introns are strongly enriched in splicing regulatory elements (Yeo et al., 2005) (see below). Also consistent with this view, a recent analysis of AS microarray data has revealed that conserved AS events undergo differential regulation between tissues more often than AS events that are species specific (Xing and Lee, 2005b). Together, these observations suggest that species-specific splice variants, relative to conserved splice variants, may less frequently play important functional roles.

Surprisingly, over one-third of AS events identified in sequenced transcripts from human and mouse introduce in-frame premature termination codons (PTCs), which have the potential to trigger transcript degradation by the process of nonsense-mediated mRNA decay (NMD) (Lewis et al., 2003). Based on this observation and experimental studies providing evidence that some protein factors, including splicing factors that bind RNA, can autoregulate their levels by binding to their own pre-mRNA to promote increased levels of PTC-containing splice variants (Soergel et al., 2005), it had been speculated that AS-coupled NMD could provide a frequent mechanism for the regulation of gene expression (Lewis et al., 2003). However, more recent analyses have shown that most PTC-containing splice variants are not conserved between human and mouse (Baek and Green, 2005; Pan et al., 2006). Moreover, a recent microarray analysis of AS events in mammalian cells and tissues suggested that most PTC-containing splice variants are produced at low levels independent of the action of NMD and are rarely subject to tissue-specific regulation (Pan et al., 2006). These results support the view that, although AS-coupled NMD may not play a widespread role in gene regulation, this process may serve to fine tune the levels of specific classes of factors, including subsets of splicing factors and other RNA binding proteins.

AS in Genome Evolution

The observations summarized above, in addition to other studies that have focused on comparisons of transcript diversity between species, reveal a fascinating aspect of AS that has undoubtedly had a major impact on genome evolution. The relatively high frequency of species-specific AS events that generate low-abundance splice variants provides a powerful "test bed" to evolve new or modified gene functions. The "minor" splice variants, which are found in low abundance, may not often have a major impact on physiology, whereas the more abundant splice variants are more likely to maintain critical gene activities. The minor splice variants therefore are relatively unconstrained and may evolve more rapidly such that they either are lost or, over a longer evolutionary time span, are subject to positive selection if the splice variant confers a selective advantage. Consistent with this view, analyses of substitution rates in the genomes of different mammalian species indicate that the exons of species-specific splice variants evolve more rapidly than constitutive or conserved alternatively spliced exons (Xing and Lee, 2005a).

It is also interesting to consider the possibility that the accumulation of low-abundance splice variants produced from a large fraction of mammalian genes (as well as noncoding RNA transcribed from intronic and intergenic regions) could have a more subtle impact on physiology and perhaps indirectly influence species-specific characteristics. For example, it is possible that many of these splice variants produce low levels of proteins. Alternatively, these splice variants may sequester nucleic acid binding proteins, reducing their availability to function in the regulation of gene expression. In analogy to the latter situation, the expansion of CUG trinucleotide repeats in the 3′UTR of the *Dmpk* gene, which is associated with myotonic dystrophy, is thought to result in the sequestration of the tissue-specific Muscleblind (MBNL) splicing factors, leading to aberrant AS patterns of disease-relevant genes (Kanadia et al., 2003). Currently, the extent to which the composition of the transcriptome itself might indirectly influence AS or other steps in gene expression by sequestration of factors is not known.

AS Networks

A major goal of current and future AS research is to systematically elucidate the functional roles of physiologically relevant AS events, as well as to establish how AS events are coordinated with other steps in gene expression to control specific cellular activities. Based largely on predictions from sequence analyses, only ~10% of AS events have the potential to delete, insert, or modify functionally defined domains in proteins, and these events provide a relatively straightforward means to design appropriate experiments to address the functional consequences of AS (Resch et al., 2004b; Pan et al., 2005). However, the remaining ~90% of AS events located in coding regions do not coincide with boundaries of functionally defined protein domains or regions that are obviously important for overall protein folding. Based on mapping of coding regions of alternative exons onto solved structures, it appears that most AS events affect coiled or "loop" regions of secondary structures, and the majority of these coding sequences are located on the surfaces of proteins (Wang et al., 2005). These regions are candidates for mediating protein-protein interactions or interactions with other types of ligands. Consistent with the possibility that AS has an important role in the formation and regulation of protein-protein interactions is the observation that, relative to transcripts of multicopy genes, transcripts of single-copy genes more often undergo AS (Kopelman et al., 2005; Su et al., 2006) and also more often encode proteins that are found at highly connected nodes in interaction networks (Hughes and Friedman, 2005). These findings emphasize the importance of employing high-throughput approaches to identify the interaction targets of specific coding regions specified by alternative exons and to integrate this information into existing knowledge of gene and protein interaction networks.

The results of several recent microarray-based analyses of AS are beginning to provide fascinating insights into how AS functions on a global level. Previous analyses using conventional microarrays revealed that groups of genes that are coregulated at the transcript level across different cell and tissue types, or under different growth conditions, tend to function in the same processes and pathways (Miki et al., 2001; Wu et al., 2002; Zhang et al., 2004). Indeed, this property of transcriptionally coregulated genes has been used to successfully predict the functions of unannotated genes (Wu et al., 2002; Zhang et al., 2004). Remarkably, sets of genes that are specifically regulated at the AS level in different cells and tissues do not extensively overlap with sets of genes regulated at the transcriptional level in the same cells and tissues (Le et al., 2004; Pan et al., 2004, 2006). This observation indicates that different subsets of genes may be coordinated at the levels of AS and transcription to regulate and maintain cell- and tissue-type-specific functions. However, the microarray data sets surveying AS in different cell and tissue types have not yet been sufficiently large to determine whether genes coregulated at the level of AS are significantly associated with common functions, as observed for transcriptionally coregulated genes.

Nevertheless, previous studies on the *Drosophila* AS regulator Transformer-2 (*Tra2*), as well as a recent study employing AS microarray analyses of tissues from mice carrying a deletion in the gene for the brain-specific factor Nova-2, have indicated that at least some regulated AS events may be under coordinated control. Tra2 regulates AS events that play an important role in controlling the function of two transcription factors, Doublesex and Fruitless, which regulate sets of genes involved in sex determination and courtship behavior, respectively (Forch and Valcarcel, 2003; Dulac, 2005). Genes containing alternative exons that undergo a change in level in the brains of mice that lack Nova-2, compared to the brains of wild-type mice, are significantly enriched in Gene Ontology terms and manually curated annotations associated with functions in the synapse and axon guidance (Ule et al., 2005). These results corroborated earlier findings by the same authors demonstrating that transcripts bound by Nova splicing factors in vivo also tend to be predominately associated with these neuronal functions (Ule et al., 2003). Moreover, many of the protein products of the genes in the Nova-2-regulated group defined by the microarray analysis have been shown in earlier studies to physically interact (Ule et al., 2005). Consistent with the earlier AS profiling studies, which indicated that the groups of genes that are regulated by AS and transcription are primarily separate, the absence of Nova-2 did not appear to have a significant effect on the transcript levels of genes monitored on the AS microarray (Ule et al., 2005).

Further studies will be required to establish to what extent alternative exons in functionally associated genes are coordinated in the form of "exon networks" that

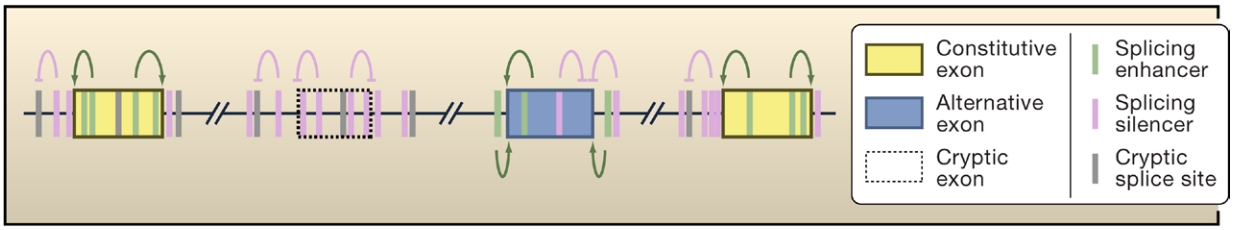

Figure 3. Elements of a "Splicing Code"
The diagram illustrates approximate relative distributions of splicing enhancer and silencer sequences in pre-mRNA in relation to constitutive, alternative, and pseudoexons, as well as to cryptic splice sites. Green arrows illustrate the positive activity of splicing enhancers (green bands) on the selection of adjacent splice sites in the alternative exon (blue box) and constitutive exons (yellow boxes). Pink arrows with flat heads indicate the negative activity of splicing silencers (pink bands) in pseudoexons and intronic regions proximal to cryptic splice sites, as well as in the alternative exon. The density of cryptic splice sites is underrepresented in the diagram for the sake of clarity.

operate in parallel with networks of genes controlled at the level of transcription and other gene regulatory processes. If this concept were found to apply on a wider scale, an important outcome would be the potential to predict the functions of thousands of currently unannotated genes on the basis of their correlated AS levels. This might be especially useful in situations where such unannotated genes do not display significantly correlated patterns of expression at the transcript level. Moreover, the ability to further detect and characterize the functions of regulated AS events on a large scale will provide an entirely new perspective on how different cellular processes are regulated and communicate with one another. The development of siRNA screens for the functional characterization of splice variants, together with high-throughput "exon-resolution" protein-protein interaction screens, has considerable potential to uncover AS networks. Ultimately, a major goal will be to integrate this information into knowledge of coordinated gene-expression programs and networks that may be controlled at other levels in addition to transcription, including "mRNP operons" (i.e., functionally coordinated sets of transcripts that are regulated at the levels of translation, stability, and export by RNA binding proteins) (Keene and Lager, 2005), miRNA-mediated regulation, and possibly other co-/posttranscriptional and translation-control mechanisms.

Defining a Sequence Code for Regulated AS

In contrast to genes in budding yeast, which contain relatively few (~200 total) and short introns that are defined by highly conserved core splicing signals, the average human gene contains 8 to 9 introns demarcated by core splicing signals that are generally poorly conserved. These core signals lack sufficient information content for the splicing machinery to distinguish correct pairs of splice sites from cryptic splice sites, which are vastly more abundant than correct splice sites (Senapathy et al., 1990; Sun and Chasin, 2000). A combination of studies employing site-directed mutagenesis of pre-mRNA reporters, binding and functional SELEX of RNA binding splicing factors, and computational strategies have resulted in the identification of additional *cis*-acting sequence elements in exons and introns that are important for promoting splice-site recognition and discriminating correct from incorrect pairs of splice sites (Cartegni et al., 2002; Matlin et al., 2005) (Figure 3). These sequences, referred to as exonic splicing enhancers/silencers (ESEs/ESSs) and intronic splicing enhancers/silencers (ISEs/ISSs), also play important roles in the regulation of AS. An important goal toward defining the functions of AS events is also to understand the mechanisms by which these *cis*-acting sequences in pre-mRNA combine with each other and with *trans*-acting splicing factors to define global splicing patterns. The availability of extensive transcript sequence data and AS profiling technologies is beginning to yield new insights into the nature of this "splicing code."

Previously detected enhancer and silencer sequences tend to be short (typically ~5–10 nt in length) and consist of relatively degenerate consensus sequences that are recognized by the RNA binding domains of a diverse array of factors at the earliest stages of spliceosome formation (Matlin et al., 2005). These factors include members of the SR and hnRNP families of proteins, which can activate and repress the formation of interaction networks across exons and introns that participate in spliceosome assembly (refer to Figure 4 for further information). The degenerate nature of enhancer and silencer elements probably reflects evolutionary selection pressure to provide flexible recognition by cognate RNA binding domains because the overlapping regions are often under selection pressure to maintain protein coding sequences.

Density distribution plots of ESE and ESS elements, as defined by both experimental and computational methods, have provided several surprising observations. In addition to being concentrated in pseudoexons and regulated alternative exons relative to constitutive exons, silencer motifs are also enriched in the intronic flanks proximal to constitutively spliced exons. This suggests that these elements could perform an important role in the precise definition of constitutive splice sites, in addition to suppressing cryptic exons and regulating AS (Zhang and Chasin, 2004; Wang et al., 2004). Moreover, recent work has also shown that diverse silencer

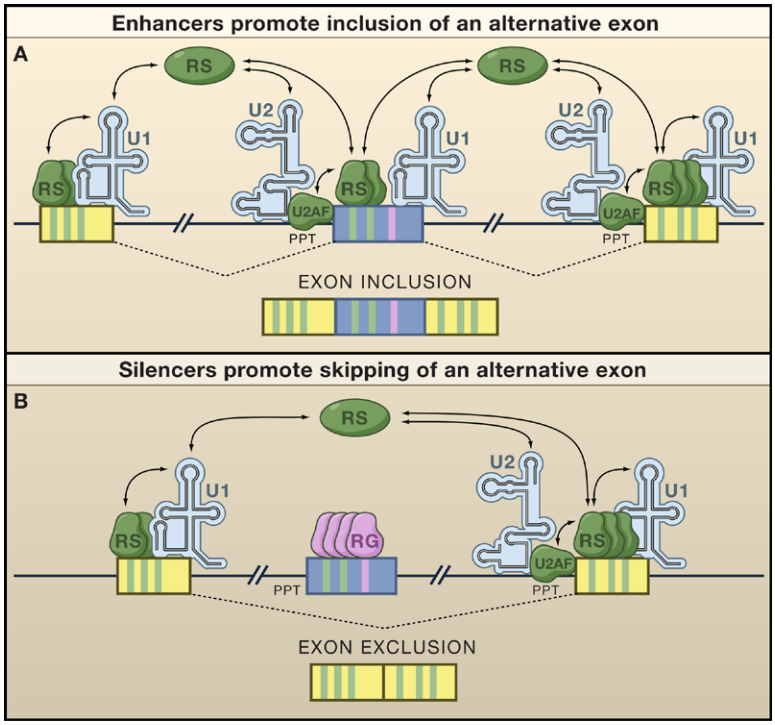

Figure 4. Regulation of Crossintron and Crossexon Interactions by SR and hnRNP Proteins

U1 and U2 small nuclear ribonucleoprotein particles (snRNPs) associate with the 5′ splice site and branch site, respectively. Binding of U2 snRNP to the branch site is promoted by the binding of an alternating arginine/serine (RS) domain-containing factor (U2AF) to the polypyrimidine tract (PPT). Binding of other RS domain proteins, such as members of the SR family and SR-related proteins, to exonic splicing enhancer (ESE) motifs (green bands) within the alternative exon (blue box) also facilitates the stable assembly of U1 and U2 snRNPs. The large RS domain protein (green oval) represents an SR-related splicing coactivator protein that serves to bridge crossintron, and possibly crossexon, interactions involving snRNPs and ESE bound SR proteins.

(A) Model for RS domain proteins in mediating the ESE-dependent inclusion of the alternative exon.

(B) Model for exonic splicing silencer (ESS) dependent skipping of the alternative exon promoted by the binding of an hnRNP protein to an ESS. Binding of the hnRNP protein to the ESS leads to its multimerization along the alternative exon, as mediated by the arginine/glycine (RG) repeat region of the protein. Multimerization of the hnRNP protein is proposed to disrupt the binding of one or more adjacent SR proteins, resulting in exon skipping. Other proposed models for competition between SR and hnRNP proteins include a steric hindrance mechanism, in which the binding of an hnRNP protein to an ESS directly displaces an adjacent SR protein, and a "looping-out" mechanism, in which binding of hnRNP proteins to distal sites within the introns flanking an alternative exon result in preferential splicing of the distal splice sites and skipping of the alternative exon (Martinez-Contreras et al., 2006). These models are not mutually exclusive and may operate in different pre-mRNAs. Also not shown are interactions involving intronic splicing enhancers (ISE) and silencers (ISS), which can function to promote or repress interactions required for the inclusion of adjacent alternative exons (see Figure 3).

sequences, as well as some ESEs, play an important role in controlling the selection of alternative 5′ and 3′ splice sites and that a specific class of silencers may also function to regulate intron retention events (Wang et al., 2006).

The combination of data on splice-site strength and the distributions of enhancer and silencer elements has helped formulate an initial "draft" splicing code that can simulate splice-site selection with considerable accuracy (Wang et al., 2004). However, what is currently missing from this draft is information allowing prediction of cell- and tissue-specific AS, as well as information allowing the accurate prediction of which protein factors are most likely to target which exons. Nevertheless, the application of siRNA-mediated knockdowns and conventional knockouts of splicing factors, together with AS microarray analyses, may provide the data sets required to at least partially bridge the current gap in information. For example, a microarray containing probes for all annotated *Drosophila* splice variants was used to analyze splicing patterns in cultured *Drosophila* S2 cells following the individual knockdown of four different proteins known to regulate AS, dASF/SF2, B52/SRp55, hrp48, and PSI (Blanchette et al., 2005). The number of AS events regulated by each factor was found to be highly variable, and some events were regulated by multiple factors. The consensus binding site for B52, a *Drosophila* SR protein defined in a previous in vitro selection assay, was found to be enriched in the exons regulated by this factor. Another AS microarray profiling study performed in mouse tissues revealed the existence of new motifs that are correlated with brain- and muscle-specific AS events (Sugnet et al., 2006). Interestingly, these motifs, as well as several other motifs correlated with brain- and muscle-specific AS as revealed by sequence-based analyses, appear to consistently reside in the intron regions proximal to regulated alternative exons (Brudno et al., 2001; Minovitsky et al., 2005; Sugnet et al., 2006). Clearly, an important goal of future studies will be to further define cell- and tissue-specific splicing regulatory motifs and how they function in conjunction with each other as well as with the more generally utilized enhancer and silencer motifs. Ultimately, such studies should facilitate formulating an expanded splicing code that will allow the prediction of cell- and tissue-specific splicing patterns.

Regulation of Splicing and Human Disease

An important outcome of many of the recent global analyses of splicing is our increased understanding of the role of splicing defects in human disease. Significantly, a more detailed knowledge of the identity and distribution of splicing enhancer and silencer sequences in exons and introns has already provided valuable information

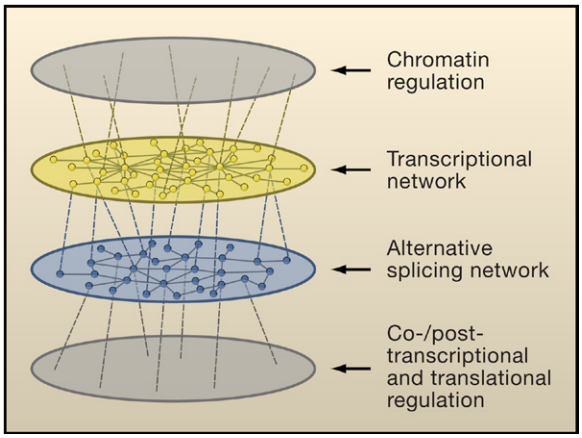

Figure 5. Layers of Coordinated Gene Regulation
Recent microarray-based studies have provided evidence that different subsets of genes are regulated at the levels of transcription and AS to define cell- and tissue-type-specific gene-expression profiles. An emerging model is that these subsets of genes may comprise "layers" of gene networks that coordinate specific cellular functions. The dotted lines indicate hypothetical interconnections that serve to integrate these and the other layers of gene regulation shown.

for improving our ability to predict which disease mutations have the potential to disrupt splicing (Cartegni et al., 2002; Fairbrother et al., 2002; Pagani and Baralle, 2004; Wang et al., 2004; Zhang and Chasin, 2004). This is important because many disease mutations that target nonsynonymous amino acid codon positions are often assumed to cause disease through changes to the amino acid sequence of a protein, whereas careful examination of transcripts from many disease genes is revealing that many such mutations can cause more serious defects by disrupting the inclusion of an entire exon or activation of a cryptic splice site. Either event has the potential to introduce a PTC resulting in degradation of the entire transcript by NMD (Cartegni et al., 2002; Pagani and Baralle, 2004). Likewise, many disease-associated mutations that target synonymous codon positions affect splicing. Examples of human disease genes known to be targeted by synonymous and nonsynonymous mutations that often affect exonic splicing control elements include *BRCA1* (breast cancer 1, early onset), *CFTR* (cystic fibrosis transmembrane conductance regulator), *HPRT1* (hypoxanthine phosphoribosyltransferase 1), *MAPT* (microtubule-associated protein tau), and *SMN1* (survival of motor neuron 1) (Cartegni et al., 2002). Such studies, as well as data stemming from several different types of analyses, including nucleotide substitution rates in the orthologous exons of different species, surveys of the complexity and distributions of ESE and ISE sequences, and more systematic analyses of transcripts from specific disease genes such as *NF1* (neurofibromatosis type 1) and *ATM* (ataxia telangiectasia mutated) (Teraoka et al., 1999; Ars et al., 2000), lead to the striking conclusion that as many as 50% of disease mutations in exons may impact on splicing. This conclusion strongly attests to the importance of introducing routine transcript analyses in order to properly assess possible mechanisms accounting for human diseases. Moreover, the high degree of conservation of intron sequences flanking conserved AS events indicates that these regions in disease genes should also be routinely examined for mutations that alter splicing.

Microarray profiling technologies present major possibilities for the future transformation of clinical practice because gene-expression signatures can provide valuable diagnostic and prognostic information, which in turn facilitates choices for appropriate treatments. The recent advent of AS microarrays and the resulting data indicate that it will be important to obtain AS signatures because profiling at the transcript level, or at other levels of gene expression, may entirely miss important disease-relevant changes in gene expression at the level of splicing. Indeed, given the recent evidence summarized above for the increased role of disease mutations in splicing, it is quite possible that microarrays enabling profiling of splicing may in many cases provide more informative diagnostic or prognostic signatures than can be obtained from conventional microarrays for profiling levels of gene expression. In support of this view, a very recent analysis of microarray data from profiling both AS and transcript levels in prostate tissues revealed that detection of splice variants can permit a more reliable discrimination of tumor from normal tissue than the detection of transcript-level differences in the same tissue samples (Li et al., 2006; Zhang et al., 2006).

Conclusions and Future Directions

Initial computational analyses of transcript sequences and microarray profiling of splicing have provided a more detailed understanding of the extent and complexity of AS in different cells and tissues. However, the question remains as to whether we are just observing the tip of the iceberg or whether the majority of important AS events have already been identified. A satisfactory answer to this question will require more extensive exon coverage by both sequencing and AS profiling strategies in a wider range of cells and tissues, including those that are less abundant and/or more specialized. Also important will be an expansion of the application and sensitivity of high-throughput methods, such as mass spectrometry and antibody microarrays, to detect splice variants at the protein level. Moreover, the range of species examined by these AS profiling methods needs to be expanded. For example, due to the lack of substantial transcript sequence data, we currently know very little about how splicing patterns compare in the chimpanzee, our closest primate relative. Yet as a major step in gene regulation, we can anticipate that differences in AS, like transcription, may play an important role in accounting for primate-species-specific characteristics.

Another important outcome of recent studies on AS is the finding that the majority of genes regulated in a tissue-specific manner by AS are different from those reg-

ulated in a tissue-specific manner at the transcriptional level. Moreover, it is emerging that groups of tissue-specific AS events may function in a coordinated manner in specific pathways or interaction networks, in much the same way as has been observed for groups of genes coregulated at the transcriptional level. This promotes the concept of "layers" of gene regulation and raises the interesting question as to the extent and nature of exon networks that may serve in parallel with other layers to coordinate gene activities and interactions so as to refine and/or expand cell- and tissue-type-specific functions (Figure 5). Other parallel layers of regulation that presumably also occur on a frequent basis include miRNA- and non-miRNA-mediated control of mRNA turnover or translation, mRNA localization, as well as other posttranscriptional and co-/posttranslational mechanisms. An important goal in the future will be to understand how these layers of regulation are integrated with one another. Clearly, understanding gene-expression regulation at an exon-level resolution will be a major goal of future research in what may well be defined as the "exonomics" era.

ACKNOWLEDGMENTS

My sincere apologies to researchers whose primary articles could not be cited due to length constraints of this review. I am grateful to my colleagues Jim Ingles, Jim Friesen, Brendan Frey, Javier Caceres, Yoseph Barash, Leo Lee, and Yi Xing, as well as to members of my laboratory including Isabel Aznarez, John Calarco, Matthew Fagnani, Joanna Ip, May Khanna, and Arneet Saltzman, for helpful discussions and comments on the manuscript. Research in my group is supported by grants from the Canadian Institutes of Health Research, the Canada Foundation for Innovation, the National Cancer Institute of Canada, and Genome Canada.

REFERENCES

Ars, E., Serra, E., Garcia, J., Kruyer, H., Gaona, A., Lazaro, C., and Estivill, X. (2000). Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Hum. Mol. Genet. *9*, 237–247.

Baek, D., and Green, P. (2005). Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. Proc. Natl. Acad. Sci. USA *102*, 12813–12818.

Beffert, U., Weeber, E.J., Durudas, A., Qiu, S., Masiulis, I., Sweatt, J.D., Li, W.P., Adelmann, G., Frotscher, M., Hammer, R.E., and Herz, J. (2005). Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2. Neuron *47*, 567–579.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Biol. *17*, 251–256.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. *72*, 291–336.

Blanchette, M., Green, R.E., Brenner, S.E., and Rio, D.C. (2005). Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. Genes Dev. *19*, 1306–1314.

Boucher, L., Ouzounis, C.A., Enright, A.J., and Blencowe, B.J. (2001). A genome-wide survey of RS domain proteins. RNA *7*, 1693–1701.

Brett, D., Pospisil, H., Valcarcel, J., Reich, J., and Bork, P. (2002). Alternative splicing and genome complexity. Nat. Genet. *30*, 29–30.

Brudno, M., Gelfand, M.S., Spengler, S., Zorn, M., Dubchak, I., and Conboy, J.G. (2001). Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. Nucleic Acids Res. *29*, 2338–2348.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. *3*, 285–298.

Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Puettmann-Holgado, R., Lamar, D.R., and Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila. Cell *125*, 607–620.

Chern, T.M., van Nimwegen, E., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Zavolan, M. (2006). A simple physical model predicts small exon length variations. PLoS Genet. *2*, e45.

Davis, C.A., Grate, L., Spingola, M., and Ares, M., Jr. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res. *28*, 1700–1706.

Dulac, C. (2005). Sex and the single splice. Cell *121*, 664–666.

Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. Mol. Cell *21*, 849–859.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007–1013.

Forch, P., and Valcarcel, J. (2003). Splicing regulation in Drosophila sex determination. Prog. Mol. Subcell. Biol. *31*, 127–151.

Graveley, B.R. (2005). Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell *123*, 65–73.

Gromak, N., Rideau, A., Southby, J., Scadden, A.D., Gooding, C., Huttelmaier, S., Singer, R.H., and Smith, C.W. (2003). The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. EMBO J. *22*, 6356–6364.

Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., Backofen, R., and Platzer, M. (2004). Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nat. Genet. *36*, 1255–1257.

Horiuchi, T., and Aigaki, T. (2006). Alternative trans-splicing: a novel mode of pre-mRNA processing. Biol. Cell. *98*, 135–140.

Hughes, A.L., and Friedman, R. (2005). Gene duplication and the properties of biological networks. J. Mol. Evol. *61*, 758–764.

Ibrahim el, C., Schaal, T.D., Hertel, K.J., Reed, R., and Maniatis, T. (2005). Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc. Natl. Acad. Sci. USA *102*, 5002–5007.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science *302*, 2141–2144.

Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W., and Swanson, M.S. (2003). A muscleblind knockout model for myotonic dystrophy. Science *302*, 1978–1980.

Keene, J.D., and Lager, P.J. (2005). Post-transcriptional operons and regulons co-ordinating gene expression. Chromosome Res. *13*, 327–337.

Kim, H., Klein, R., Majewski, J., and Ott, J. (2004). Estimating rates of alternative splicing in mammals and invertebrates. Nat. Genet. *36*, 915–916; author reply 916–917.

Kopelman, N.M., Lancet, D., and Yanai, I. (2005). Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat. Genet. *37*, 588–589.

Kornblihtt, A.R. (2005). Promoter usage and alternative splicing. Curr. Opin. Cell Biol. *17*, 262–268.

Kornblihtt, A.R. (2006). Chromatin, transcript elongation and alternative splicing. Nat. Struct. Mol. Biol. *13*, 5–7.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Le, K., Mitsouras, K., Roy, M., Wang, Q., Xu, Q., Nelson, S.F., and Lee, C. (2004). Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. Nucleic Acids Res. *32*, e180.

Lee, C., Atanelov, L., Modrek, B., and Xing, Y. (2003). ASAP: the Alternative Splicing Annotation Project. Nucleic Acids Res. *31*, 101–105.

Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. (2003). The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. Science *300*, 1288–1291.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc. Natl. Acad. Sci. USA *100*, 189–192.

Li, H.R., Wang-Rodriguez, J., Nair, T.M., Yeakley, J.M., Kwon, Y.S., Bibikova, M., Zheng, C., Zhou, L., Zhang, K., Downs, T., et al. (2006). Two-dimensional transcriptome profiling: identification of messenger RNA isoform signatures in prostate cancer from archived paraffin-embedded cancer specimens. Cancer Res. *66*, 4079–4088.

Lipscombe, D. (2005). Neuronal proteins custom designed by alternative splicing. Curr. Opin. Neurobiol. *15*, 358–363.

Martinez-Contreras, R., Fisette, J.F., Nasim, F.U., Madden, R., Cordeau, M., and Chabot, B. (2006). Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol. *4*, e21.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. *6*, 386–398.

Miki, R., Kadota, K., Bono, H., Mizuno, Y., Tomaru, Y., Carninci, P., Itoh, M., Shibata, K., Kawai, J., Konno, H., et al. (2001). Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. Proc. Natl. Acad. Sci. USA *98*, 2199–2204.

Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I., and Conboy, J.G. (2005). The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. Nucleic Acids Res. *33*, 714–724.

Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. Nat. Genet. *30*, 13–19.

Modrek, B., and Lee, C.J. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat. Genet. *34*, 177–180.

Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res. *29*, 2850–2859.

Ohler, U., Shomron, N., and Burge, C.B. (2005). Recognition of unknown conserved alternatively spliced exons. PLoS Comput. Biol. *1*, 113–122.

Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. Nat. Rev. Genet. *5*, 389–396.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., et al. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol. Cell *16*, 929–941.

Pan, Q., Bakowski, M.A., Morris, Q., Zhang, W., Frey, B.J., Hughes, T.R., and Blencowe, B.J. (2005). Alternative splicing of conserved exons is frequently species-specific in human and mouse. Trends Genet. *21*, 73–77.

Pan, Q., Saltzman, A.L., Kim, Y.K., Misquitta, C., Shai, O., Maquat, L.E., Frey, B.J., and Blencowe, B.J. (2006). Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. Genes Dev. *20*, 153–158.

Philipps, D.L., Park, J.W., and Graveley, B.R. (2004). A computational and experimental approach toward a priori identification of alternatively spliced exons. RNA *10*, 1838–1844.

Resch, A., Xing, Y., Alekseyenko, A., Modrek, B., and Lee, C. (2004a). Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. Nucleic Acids Res. *32*, 1261–1269.

Resch, A., Xing, Y., Modrek, B., Gorlick, M., Riley, R., and Lee, C. (2004b). Assessing the impact of alternative splicing on domain interactions in the human proteome. J. Proteome Res. *3*, 76–83.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell *101*, 671–684.

Senapathy, P., Shapiro, M.B., and Harris, N.L. (1990). Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol. *183*, 252–278.

Shai, O., Morris, Q.D., Blencowe, B.J., and Frey, B.J. (2006). Inferring global levels of alternative splicing isoforms using a generative model of microarray data. Bioinformatics *22*, 606–613.

Soergel, D.A.W., Lareau, L.F., and Brenner, S.E. (2005). Regulation of gene expression by coupling of alternative splicing and NMD. In Nonsense-Mediated mRNA Decay, L.E. Maquat, ed. (Georgetown, TX: Landes Bioscience), pp. 175–196.

Sorek, R., Shamir, R., and Ast, G. (2004). How prevalent is functional alternative splicing in the human genome? Trends Genet. *20*, 68–71.

Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., et al. (2004). A gene expression map for the euchromatic genome of Drosophila melanogaster. Science *306*, 655–660.

Su, Z., Wang, J., Yu, J., Huang, X., and Gu, X. (2006). Evolution of alternative splicing after gene duplication. Genome Res. *16*, 182–189.

Sugnet, C.W., Srinivasan, K., Clark, T.A., O'Brien, G., Cline, M.S., Wang, H., Williams, A., Kulp, D., Blume, J.E., Haussler, D., and Ares, M. (2006). Unusual intron conservation near tissue-regulated exons found by splicing microarrays. PLoS Comput. Biol. *2*, e4.

Sun, H., and Chasin, L.A. (2000). Multiple splicing defects in an intronic false exon. Mol. Cell. Biol. *20*, 6414–6425.

Teraoka, S.N., Telatar, M., Becker-Catania, S., Liang, T., Onengut, S., Tolun, A., Chessa, L., Sanal, O., Bernatowska, E., Gatti, R.A., and Concannon, P. (1999). Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am. J. Hum. Genet. *64*, 1617–1631.

Thanaraj, T.A., Stamm, S., Clark, F., Riethoven, J.J., Le Texier, V., and Muilu, J. (2004). ASD: the Alternative Splicing Database. Nucleic Acids Res. *32*, D64–D69.

Ule, J., and Darnell, R.B. (2006). RNA binding proteins and the regulation of neuronal synaptic plasticity. Curr. Opin. Neurobiol. *16*, 102–110.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat. Genet. *37*, 844–852.

Wang, H., Hubbell, E., Hu, J.S., Mei, G., Cline, M., Lu, G., Clark, T., Siani-Rose, M.A., Ares, M., Kulp, D.C., and Haussler, D. (2003). Gene structure-based splice variant deconvolution using a microarray platform. Bioinformatics *19* (*Suppl 1*), i315–i322.

Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831–845.

Wang, P., Yan, B., Guo, J.T., Hicks, C., and Xu, Y. (2005). Structural genomics analysis of alternative splicing and application to isoform structure modeling. Proc. Natl. Acad. Sci. USA *102*, 18920–18925.

Wang, Z., Xiao, X., Van Nostrand, E., and Burge, C.B. (2006). General and specific functions of exonic splicing silencers in splicing control. Mol. Cell *23*, 61–70.

Watahiki, A., Waki, K., Hayatsu, N., Shiraki, T., Kondo, S., Nakamura, M., Sasaki, D., Arakawa, T., Kawai, J., Harbers, M., et al. (2004). Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas. Nat. Methods *1*, 233–239.

Watson, F.L., Puttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., and Schmucker, D. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science *309*, 1874–1878.

Wu, L.F., Hughes, T.R., Davierwala, A.P., Robinson, M.D., Stoughton, R., and Altschuler, S.J. (2002). Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat. Genet. *31*, 255–265.

Xing, Y., and Lee, C. (2005a). Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc. Natl. Acad. Sci. USA *102*, 13526–13531.

Xing, Y., and Lee, C.J. (2005b). Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. PLoS Genet. *1*, e34.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., and Fu, X.D. (2002). Profiling alternative splicing on fiber-optic arrays. Nat. Biotechnol. *20*, 353–358.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. Genome Biol. *5*, R74.

Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T., and Burge, C.B. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. Proc. Natl. Acad. Sci. USA *102*, 2850–2855.

Zavolan, M., Kondo, S., Schonbach, C., Adachi, J., Hume, D.A., Hayashizaki, Y., and Gaasterland, T. (2003). Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. Genome Res. *13*, 1290–1300.

Zhang, C., Li, H.R., Fan, J.B., Wang-Rodriguez, J., Downs, T., Fu, X.D., and Zhang, M.Q. (2006). Profiling alternatively spliced mRNA isoforms for prostate cancer classification. BMC Bioinformatics *7*, 202.

Zhang, W., Morris, Q.D., Chang, R., Shai, O., Bakowski, M.A., Mitsakakis, N., Mohammad, N., Robinson, M.D., Zirngibl, R., Somogyi, E., et al. (2004). The functional landscape of mouse gene expression. J. Biol. *3*, 21.

Zhang, X.H., and Chasin, L.A. (2004). Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev. *18*, 1241–1250.

Zheng, C.L., Kwon, Y.S., Li, H.R., Zhang, K., Coutinho-Mansfield, G., Yang, C., Nair, T.M., Gribskov, M., and Fu, X.D. (2005). MAASE: an alternative splicing database designed for supporting splicing microarray applications. RNA *11*, 1767–1776.